DOCUMENT RESUME

ED 079 402 TM 003 006

AUTHOR Woodson, M. I. Charles E.
TITLE The Issue of Item and Test Variance for Criterion-Referenced Tests.
NOTE 5p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Criterion Referenced Tests; *Evaluation Criteria; Item Analysis; *Item Sampling; Statistical Analysis; *Test Construction; Testing; Test Interpretation

ABSTRACT

It has been argued that item variance and test variance are not necessary characteristics for criterion-referenced tests, although they are necessary for norm-referenced tests. This position is in error because it considers sample statistics as the criteria for evaluating items and tests. Within a particular sample, an item or test may have no variance, but in the population for which the test was designed and evaluated, both items and tests must have variance. (Author)

FILMED FROM BEST AVAILABLE COPY

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 079402

The Issue of Item and Test Variance for Criterion-Referenced Tests*

M. T. Chas. E. Woodson[2]

University of California, Berkeley

Abstract

It has been argued that item variance and test variance are not necessary characteristics for criterion-referenced tests, although they are necessary for norm-referenced tests. This position is in error because it considers sample statistics as the criteria for evaluating items and tests. Within a particular sample, an item or test may have no variance, but in the population for which the test was designed and evaluated, both items and tests must have variance.

TM 003 006

The Issue of Item and Test Variance for Criterion-Referenced Tests[1]

M. I. Chas. E. Woodson

University of California, Berkeley

Popham and Husek (1969) have argued that the test items for criterion-referenced tests and the tests themselves (Popham, personal communication) may have no variance and therefore traditional methods of empirical evaluation of test items and tests are invalid for criterion-referenced tests. Popham and Husek (1969) conclude, "With criterion-referenced tests, variability is irrelevant . . . . Variability is not a necessary condition for a good criterion-referenced test."

Consider the example of the ideal outcome of a perfect instructional procedure: Before instruction everyone misses all items, after instruction everyone gets all items correct. This ideal outcome has been referred to as showing no item variance.

The basic flaw in this argument is that it fails to consider the question of what generalizations are to be made from the observations. Popham and Husek were speaking about sample statistics, but statistics for evaluating items or test characteristics refer to the population of observations for which the instrument was designed and calibrated. The population of observations for which an item is calibrated must be the reference for the evaluation of an item. In classical test theory (Lord

and Novick, 1968) for norm-referenced tests, item analysis, test development and test validation must be done on observations of a sample representative of the population for which the test will be used. For criterion-referenced tests, item analysis and test development must be done on observations representative of the observations within the range of interest on the characteristic of interest. Referring to the ideal outcome above, the range of possible observations of interest includes the observation of passing no items as well as that of passing all items.

Statistics from a particular sample of observations, perhaps restricted in some way, do not give us a definitive answer about what the discrimination in the population of observations may be. If the sample is considerably restricted, the estimates of test parameters will be influenced. Items may have no variance in a restricted sample, e.g., individuals who have finished an instructional program, and yet be useful items because they do have variance within the population for which they are calibrated. The extreme case in which an item is missed by all subjects on the pre-test and answered correctly by all subjects on the post-test, in fact, is an example of the maximum variance for an item within the sample of observations collected. In this case we have data from what appears to be two extreme points on the characteristic of interest.

In classical theory, item analysis seeks to answer the question, "Does the item discriminate on the characteristic measured within the distribution of scores for the population of interest." The reference here is to differences among persons of a population. Item analysis in this case requires observations on a sample representative of this population.

Criterion-referenced item analysis seeks an answer to the question, "Does the item discriminate within the range of interest on the characteristic measured?" The reference here is to different observations on the characteristic. Item analysis in this case requires observations at different points on the characteristic.

In either case, item variance and discrimination are essential. In short, (1) items and tests must be evaluated for the range of the characteristic for which they will be used, and (2) items and tests which give no variability in this population of observations, give no information and are therefore not useful.

References

Popham, W. J. & Husek, T. R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.

Lord, F. M. & Novick, M. R. Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley, 1968.